

**TITLE**

A match made in heaven. Using parallel corpora and multinomial logistic regression to analyze the expression of possession in Old Spanish

**NAMES & AFFILIATIONS OF AUTHORS**

Malte Rosemeyer, Albert-Ludwigs-Universität Freiburg

Andrés Enrique-Arias, Universitat de les Illes Balears & Harvard University

**TO APPEAR IN *LANGUAGE VARIATION AND CHANGE***

**ABSTRACT**

This study applies multinomial regression analysis to a parallel corpus of Spanish medieval translations of the Bible in order to study the different factors that condition variation in the expression of possession in Old Spanish. Our methodology allows us to determine the degree to which less frequent possessive constructions (ART+POSS, as in *la su casa* ‘the his house’, GEN, as in *la casa de él* ‘the house of him’ and ART/BARE, as in *la casa* ‘the house’) can be considered competitors to the dominant POSS construction (as in *su casa* ‘his house’) as a function of usage context differences. In comparison to the POSS construction, the ART+POSS construction usually expresses pragmatic functions such as reverence, the GEN construction is typically used to either disambiguate a reference and the ART/BARE construction is bound to contexts in which the possessor is highly accessible. Crucially, the analysis also sheds light on historical changes in the balance between structural and contextual constraints on the use of these different variants. Whereas in the 13th century, structural and stylistic constraints are almost equally important, the importance of structural constraints diminishes in the 15th century. The study thus illustrates how in reductive processes of language change, variation due to structural constraints yields to stylistic variation.

Keywords: quantitative linguistics, variationist linguistics, variable context, historical linguistics, multinomial regression analysis, possession, Spanish

**ACKNOWLEDGEMENTS**

The authors are grateful to Sali Tagliamonte, audiences in Oslo, Freiburg, Zurich, Naples and Cádiz, the editors of LVAC and three anonymous reviewers for helpful comments on

previous versions of this paper. Andrés Enrique-Arias conducted this research during a stay at the Department of Linguistics of Harvard University (2013-15).

## INTRODUCTION

A common assumption in quantitative studies of syntactic change is that change is the result of a competition between alternative expressions with the same discourse-pragmatic meaning. Therefore, a careful quantitative description of the development of the constraints on the selection of one syntactic variant over the other can reveal crucial information regarding the change in question. Accordingly, in order to correctly implement variationist methodology it is necessary to carefully delimit the *variable context*, identifying each construction that can serve to express the same discourse-pragmatic function before the quantitative analysis. In synchronic variationist linguistics, this definition of the variable context is often carried out by a combination of careful qualitative analysis and introspection (cf. Tagliamonte, 2006:78-83). But these methods are not always reliable in historical linguistics because (a) we can never be sure from the analysis of isolated examples that we have identified all of the constructions with a certain function, and (b) direct introspection is not available for historical data. As a result, it can be difficult for quantitative analyses of syntactic change to establish a correct definition of the variable context.

In addition, all quantitative studies of syntactic change face the problem of the comparability of contexts. In order to study diachronic change we have to make sure that the data we draw from texts of different periods are indeed in a relation of equivalence to each other and thus allow for a comparison. In order to achieve this, it would be necessary to locate and to examine a large number of occurrences of the *same* linguistic structure in versions that were produced at different time periods. Ideally, these occurrences should proceed from texts that have been influenced by the same textual conventions.

In this paper, we develop a methodology that contributes to alleviating these problems. This analytical approach relies on the use of parallel corpora coupled with multinomial logistic regression analysis. As an example, we analyze the expression of non-predicative possession in Old Spanish, and its development between the 13th and 15th centuries. The

methodology that we present incorporates a wider inventory of expression units for Old Spanish non-predicative possession rather than reducing this constructional network to a binary opposition between two members of that network.

This paper is structured as follows. In the following section we elaborate on the two key problems for quantitative approaches to syntactic changes mentioned above using non-predicative possession in Old Spanish. The ‘Analytical Approach’ section describes our data collection and annotation. Following this, we present the results and a discussion of the results from the analysis. The paper concludes with a summary of the findings and their analytical value in the last section.

#### TWO KEY PROBLEMS IN QUANTITATIVE APPROACHES TO SYNTACTIC CHANGE

##### *The problem of the definition of the variable context*

If, as mentioned above, syntactic variation in language reflects the competition between alternative expressions of the same meaning, a variationist analysis of a syntactic phenomenon presupposes the identification of this meaning. This corresponds to the identification of the *variable context*, within which the competition between the alternative expressions takes place. According to the *principle of accountability* (Labov, 1982:30), a quantitative analysis of the variation between different means of expression has to take into account all occurrences of this target variable. Thus, “the occurrence of variants can be calculated out of the total number of contexts in which it could have occurred, but did not” (Tagliamonte, 2006:72).

This approach creates the obvious problem of whether two types of linguistic expressions can really have the same meaning. According to the criterion of *weak complementarity*, two variants in a variable context do not need to express exactly the same meaning, i.e., have semantic equivalence. Rather, these variants merely need to express similar discourse functions, i.e., have discourse equivalence (Sankoff & Thibault, 1981:207).

Although this criterion offers a remedy to the theoretical problem of equivalence, it does not always eliminate the practical problems with the selection of the variable context. In particular, it is frequently difficult to determine which linguistic elements count as variants with discourse equivalence.

As an example, consider the expression of non-predicative possession in Old Spanish. Most research has focused on the variation in the use of the definite article preceding the possessive marker (*la mi casa* ‘the my house’, henceforth ART+POSS), a structure that is absent from present-day standard Spanish, as opposed to possessive alone (*mi casa* ‘my house’, henceforth POSS). A first research tradition has focused on stylistic factors. It has been suggested that, because ART+POSS is a structure that emphasizes possession, it is used with stylistic functions such as expressivity, solemnity, poeticality or reverence (Eberenz, 2000:265-319; Lapesa, 2000 [1970]). A second research tradition focuses on the influence of structural factors; for instance, Wanner (2005:39-40) points out that the first and second person, or singular possessors, as well as possessive structures embedded in prepositional phrases, favor the use of the ART+POSS construction. Finally, syntax-discourse factors have been taken into consideration: Company Company (2006) claims that ART+POSS is relatively more frequent when the possessor or the possessed referent has been mentioned in the previous discourse, or when the possessed referent has a high degree of accessibility by virtue of a relationship of inherent possession.

Crucially, however, there are other constructions that also serve to express possession in Old Spanish, such as a genitive phrase with a personal pronoun (*la casa de él* ‘the house of him’, henceforth GEN), the *strong possessive adjective construction*, in which the possessive is postposed (*la casa suya* ‘the house his’), or even a simple *determiner + noun construction* (ART/BARE) as in *levantó la mano* ‘he raised the [=his] hand.’ All of these structures can appear in contexts similar to those of the ART+POSS and POSS structures and their appearance correlates with a complex set of structural and contextual factors such as lyrical

register, ambiguity of reference, or cognitive prominence (Company Company, 1994; Eberenz, 2000:299; Enrique-Arias, 2012a:827-828).

The wide range of constructions with a possible possessive function illustrates an obvious analytical problem with the identification of the variable context. *Prima facie*, it is impossible to know which of these constructions fulfill the criterion of weak complementarity. A comparison of the use of, e.g., the ART+POSS construction with the use of the POSS alone presupposes that these two means of expressing possession in Old Spanish have a stronger functional similarity with each other than with the other constructional types. Such a restriction to a subset of constructional types, however, might lead to an overgeneralization of the effects that are encountered in the analysis. For instance, some studies suggest that the parameter ‘inherent possession’ (i.e., the degree of conceptual proximity between the possessor referent and possessed referent) influences the alternation between ART+POSS and POSS and the alternation between these two constructions and ART/BARE. In other words, at least with regard to this criterion, one could argue that the two variants *la su mano* ‘the his hand’ (ART+POSS) and *la mano* ‘the hand’ (ART/BARE) are more similar to each other than, e.g., *la su mano* and *su mano* ‘his hand’ (POSS). Crucially, this would mean that existing analyses that do not take into account all these constructions might miss a piece of the puzzle, as the commonality between the constructions might be a criterion of its own.

This analytical problem has consequences for diachronic analyses as well. Although the diachronic trend whereby the use of the referentially overspecified forms (ART+POSS and GEN) is eroded by Early Modern Spanish is clear, the uncertainty regarding the exact nature of the functional interdependencies between the constructional types in synchrony entails an uncertainty regarding the nature of the historical process. For instance, it is usually assumed that between Old and Early Modern Spanish, ART+POSS was replaced by POSS. However, if the degree of functional similarity of these two constructions relative to other possessive

constructions is lower than expected, it is uncertain whether it indeed was the POSS construction that replaced ART+POSS.

*The problem of the comparability of contexts*

A second important problem for the study of syntactic change concerns the comparability of contexts. It is a universal challenge to historical linguistics that historical data are always fragmentary and unrepresentative, as the selection of texts that has survived until today is the product of chance. Longitudinal analyses of syntactic change, however, need language examples that differ with regard to the state of development of the language rather than their usage contexts. This methodological challenge has been formulated in terms of a *comparability paradox* in historical corpus design (Enrique-Arias, 2012b:97): a historical corpus has to be *diverse* because it must contain texts that represent different periods, genres or dialects. At the same time, this corpus must be *uniform* in that the distribution of content type, genres or dialects along the different chronological sections in the corpus must be as similar as possible so they can be compared. In quantitative studies of syntactic change we therefore face sample-related problems concerning which texts to compare. Even if we restrict our sample to, let us say, narrative texts, we may well end up with works created under diverse textual conventions and in which the distribution of narration, description and dialogue will differ considerably. For instance, Company Company's study (2009) on the alternation of ART+POSS and POSS uses a corpus that includes epic poetry (the 12th century *Cantar de Mio Cid*), historiographical texts (such as the 13th century *General estoria*), a theatre play (the 15th century *Celestina*), a picaresque novel (the 16th century *Lazarillo de Tormes*), and other diverse texts. In such data there is no way to be certain to what extent the frequency changes attested in the analysis correspond to structural changes affecting the construction under study, and not to divergences in the setup of the data for each time period.

ANALYTICAL APPROACH



In this section we propose an analytical approach that is able to alleviate the two challenges identified in the previous section. In particular, we apply multinomial regression analysis to data taken from a parallel corpus of Bible translations in Old Spanish. This procedure allows for (a) an identification of the range of constructions used to express possession by the translators of the Bible into Old Spanish, and (b) a measurement of the historical development of these constructions in the same or very similar usage contexts over time.

*Using parallel corpora in quantitative approaches to syntactic change*

The data used in our analysis comes from *Biblia Medieval*, a parallel corpus of Old Spanish biblical translations (Enrique-Arias & Pueyo Mena, 2008-2016). The corpus includes the original texts (the Hebrew Bible and the Latin Vulgate) and their translational equivalents in medieval Spanish. The texts are aligned so that it is possible to identify the pairs or sets of sentences, phrases and words in the original text and their correspondences in the Spanish versions. The use of parallel or comparable corpora in contrastive studies (between languages, dialects, contact varieties, historical periods, etc.) today constitutes a well-established practice within both corpus linguistics (McEnery & Xiao, 2007) and sociolinguistics (Tagliamonte, 2012:162).

One of the limitations of corpus linguistics methodology is that it gravitates towards the search of explicit markers considering a closed list of elements. If we want to study the expression of possession in Old Spanish by a search in a conventional corpus, the relevant usage contexts are identified by searching for the linguistic elements instantiating these contexts, such as the possessive adjectives *mi* ‘my’, *tu* ‘your’, *su* ‘his/her/your/its’, etc. In contrast, parallel corpus methodology is much more open as we can search for *any* element used to express a given function. For instance, in a parallel corpus like *Biblia medieval* it is possible to search for possessive pronouns in the Latin Vulgate (i.e., *meus* ‘my’, *tuus* ‘your’,

*eius* ‘his/her/its’, etc.) and then observe how they are translated in the Spanish versions. As a concrete example consider the translations of Maccabees 1 5:5 in (1).<sup>1</sup>

- (1) *et incendit turres eorum*
- ‘and he burned their towers’ (MAC1 5:5)
- |    |          |                           |              |            |               |                        |
|----|----------|---------------------------|--------------|------------|---------------|------------------------|
| a. | <i>e</i> | <i>quem-ó=les</i>         |              | <i>las</i> | <i>torres</i> | (E6)                   |
|    | and      | burn-PST.PFV.3.SG=PRO.DAT |              | DET.F.PL   | towers        |                        |
| b. | <i>e</i> | <i>Puso</i>               | <i>fuego</i> | <i>a</i>   | <i>las</i>    | <i>torres</i> (General |
|    | and      | put.PST.PFV.3.SG          | fire         | to         | DET.F.PL      | towers Estoria)        |
| c. | <i>e</i> | <i>Encendió</i>           |              | <i>las</i> | <i>torres</i> | <i>d’=ellos</i> (E4)   |
|    | and      | ignite.PST.PFV.3.SG       |              | DET.F.PL   | towers        | of=them                |

Here the researcher can observe without limitations what linguistic structures are used to convey the functions expressed by *turres eorum* ‘their towers’ in this context: we find a dative pronoun and a NP with a definite article in (1a), an NP with no explicit possessive marker in (1b), and a genitive phrase in (1c).

Another problem that parallel corpus methodology helps us to overcome is that of the comparability of contexts. In parallel texts we have direct access to the evolution of linguistic structures, as translation equivalents are likely to be inserted in the same, or very similar, syntactic, semantic and pragmatic contexts of occurrence. As a parallel corpus like *Biblia Medieval* puts the discourse contextual factors largely in control, the behavior of the elements used to express possession can be observed and compared in a focused manner. Another interesting feature of the Bible is that it encompasses texts of varied textual typology: narrative, legislative, lyrical poetry, wisdom literature, epistles and dialogues. As a result, the *Biblia Medieval* corpus is particularly appropriate to explore register variation, as it is

<sup>1</sup> All passages quoted are from the *Biblia Medieval* corpus. We follow the standard practice of quoting Old Spanish biblical manuscripts from the library in the Escorial Monastery by using the letter E plus the final digit in the signature (thus Escorial I.i.6 is quoted as E6, Escorial I.i.4 is E4 and so forth). For a review of the most important issues in regards to dating, description and content of the Old Spanish biblical manuscripts contained in the corpus and for information on the abbreviations used to cite them, see [www.bibliamedieval.es](http://www.bibliamedieval.es).

possible to examine how the same translator selects language options that are appropriate for each of the genres represented in the Bible.

There are, nonetheless, some conceivable problems associated with the use of biblical translations in linguistic research. Because they are translated texts, they pose the risk of interference from the source language. As sacred texts, they may also exhibit stylistically marked language (i.e. deliberate archaisms). We must keep in mind, however, that most Spanish medieval texts in current corpora come from translations (from Latin, Arabic, Hebrew, French, etc.) or are subject to the strong Latinate influence that is characteristic of 15th century Spanish writing; moreover, non-translated secular texts such as legal documents, or literature, especially poetry, may be highly artificial as well. In sum, biblical texts are not necessarily worse sources of data relative to other medieval text types.<sup>2</sup>

At any rate, the methodological reliability of using translated texts in linguistic research crucially depends on the nature of the phenomenon to be studied. In the case at hand, variation between POSS, ART+POSS and ART/BARE in Spanish can hardly be affected by features of the original texts, as Latin has no articles and Hebrew does not employ them in possessive structures. As for GEN, while it is true that its appearance has been associated with the imitation of *eius* type structures in Latin models this does not constitute an automatic calque: it has been shown that translators use this structure selectively to remedy the referential opacity of POSS in regard to gender and number of the possessor or to exploit its stylistic possibilities (Enrique-Arias, 2012a:828).

Finally, we must acknowledge the difficulties in providing clear-cut criteria for defining what possession is. Rather than a closed list of neatly defined structures, we find a continuum of constructions that goes from morphologically marked ones to others that are only discourse-inferable and could simply have a relational meaning. In this paper we will restrict our queries to the different translations corresponding to the paradigm of Latin and Hebrew

---

<sup>2</sup> For detailed discussions of the methodological soundness of biblical texts as data sources in linguistic research see Kaiser (2005), De Vries (2007) or Enrique-Arias (2008, 2009, 2012b, 2013).

possessive pronouns. This procedure will exclude, for instance, possessive constructions with a lexical NP possessor such as *la casa de Juan* ('the house of John').

#### *Combining parallel corpus data with multinomial logistic regression analysis*

Like other regression analyses, multinomial logistic regression (henceforth MLR) analysis allows us to investigate the relationship of a set of predictor variables to a dependent variable (Orme & Combs-Orme, 2009:Ch. 3). The main areas in which these analyses have been employed are psycholinguistic studies (e.g., Li, Schweickert & Gandour, 2000; Ratcliff & McKoon, 2001) and corpus semantics (e.g., Krawczak, 2014). MLR analyses rely on the assumption of *independence of irrelevant alternatives* (IIA), which "means that your choice of Candidate A over Candidate B is not influenced by whether Candidate X joins the fray" (Orme & Combs-Orme, 2009:118). The IIA can be understood as referring to the problem of the variable context: we are only allowed to conduct a MLR analysis if we are certain that we have included all expressions that might serve to express the relevant discourse-pragmatic function and not just a subset. Because the variable context is defined on the basis of the heuristic function of translated texts, we consider the combination between parallel corpus data and multinomial logistic regression analysis to be a 'match made in heaven'. Using MLRs on parallel corpus data should allow us to model the domain of expression of a certain semantic relation much more reliably than statistical models that have a dependent variable with a number of levels corresponding to just a limited subset of possible expressions of the semantic relation.

#### *Data collection and annotation*

*Data collection.* We began the data collection procedure by entering the search string in (2) in the Latin version of the following biblical sections: *Song of Solomon* (CA), *Daniel* 1-6 (DAN), *Judges* 13-16 (JU), *Lamentations* (LA), and *Samuel I* 17 (SAM1). The criterion for this selection was to obtain language samples that reflect register variation: whereas *Daniel*,

*Judges* and *Samuel I* are narrative texts, *Song of Solomon* and *Lamentations* are lyrical texts.

This query rendered a total of 905 tokens of possessive expressions in the Latin Vulgate.

(2)     *meus | mea | meum | meam | mei | meae | meo | meos | meas | meis | meorum | tuus | tua | tuum | tui | tuae | tuorum | tuarum | tuo | tuis | tuam | tuos | tuas | tue | suus | sua | suum | sui | suae | suorum | suarum | suo | suis | suam | suos | suas | sue | nostr\* | vestr\* | eius | eorum | illius | illorum | earum | ipsius | illarum*

Then we selected all the Spanish passages that corresponded to the Latin possessives in six Old Spanish Bible manuscripts from the 13th and 15th centuries, namely the E6/8 and the *General estoria* (13th century), and the E3, E5/7, E4 and Arragel (15th century). Additionally, we conducted extensive searches of the Spanish possessive markers identified in the first step in the remaining Spanish passages. This way we added to the database a few more possessive structures that did not correspond to possessive forms in the Latin original; in all cases these corresponded with passages in which there is a possessive marker in the Hebrew version but for which the Vulgate does not employ a possessive form. Cases in which the possessive pronoun was not translated using a nominal construction in one of the translations but paraphrased in a different way were excluded from the analysis. This yielded a total of 4803 tokens. After the exclusion of three tokens of the strong possessive adjective construction (as in *la casa suya*, lit. ‘the house his’), we were left with an eventual total of 4800 cases.

*Data annotation.*     We undertook a token-by-token annotation of the Spanish data for a dependent variable (the type of possessive construction employed by the translator) and a series of predictor variables elected on the basis of the results of the previous studies summarized in the second section.

A variety of possessive constructions were used to translate the Latin possessive phrases, exemplified in (3–8). The translations are taken from the King James Bible.

- (3) *fermosas son tus quixadas* [POSS]  
 beautiful be.PRS.3PL POSS.2PL cheeks  
 ‘Thy cheeks are comely’ (CA 1:10, E3, 15th c.)
- (4) *los dientes asy como el rebaño de=las ovejas* [ART]  
 DET.M.SG teeth so like DET.M.SG flock of= DET.F.PL sheep  
 ‘Thy teeth are as a flock of sheep’ (CA 6:6, E5/7, 15th c.)
- (5) *met-ió mano a su çurrón* [BARE]  
 put-PST.PFV.3SG hand in POSS.3SG Bag  
 ‘[David] put his hand in his bag’ (SAM1 17:49, GE, 13th c.)
- (6) *Descubr-ió como la huerta la su* [ART+POSS]  
 take.away-PST.PFV.3SG like the garden DET.F.SG POSS.3SG  
*choça*  
 tabernacle  
 ‘And he hath violently taken away his tabernacle, as if it were of a garden’ (LA 2:6, E5/7, 15th c.)
- (7) *e respond-ió el padre d=ella* [DET+GEN]  
 and respond-PST.PFV.3SG DET.M.SG father of=her  
 ‘And her father said’ (JU 15:2, Arr., 15th c.)
- (8) *non=le dex-ó su padre d=ella entrar* [SU+GEN]  
 not=him let-PST.PFV.3SG POSS.3SG father of=her enter  
 ‘But her father would not suffer him to go in’ (JU 15:1, E4, 15th c.)

Most frequent are POSS and ART+POSS (Table 1). Two variants occurred with a rather low frequency and were collapsed with formally or functionally affine variants in order to obtain larger numbers. First, we only encountered 24 cases of the SU+GEN construction (8), which were combined with DET+GEN (7) in one new type, GEN, sharing a formal feature that clearly sets them apart from the others constructional types—the addition of a

prepositional phrase indicating the possessor after the possessed entity. Second, the 56 cases in which the possessive was translated with a bare noun phrase, as in (5), were assumed to behave similarly to cases in which the possessive was translated with a NP including an article, such as (4), particularly because both constructional types appear to strongly depend on a relationship of inherent possession. Consequently, we unified the constructional types BARE and ART under the new category ART/BARE.<sup>3</sup> As a result, the dependent variable TYPE has the four levels POSS (the reference level,  $n=2611$ ), ART+POSS ( $n=1747$ ), ART/BARE ( $n=307$ ), and GEN ( $n=135$ ).

TABLE 1. *Usage frequency of possessive constructions in the corpus*

POSS	ART+POS S	ART	BARE	SUYO	DET+GEN	SU+GEN	TOTAL
2611	1747	251	56	3	111	24	4803

The predictor variables, involving the factors that have been identified in previous studies, are summarized in Table 2.

---

<sup>3</sup> Results of a regression model in which ART and BARE constructions are kept apart indicate that, for both the 13th and 15th century, the influence of most predictor variables on the distribution of ART and BARE was indistinguishable.

TABLE 2. *Summary of predictor variables*

Variable name	Short description	Levels
PS_PERSON	Person morphology	1st, 2nd, 3rd
PS_NUMBER	Number morphology	Singular, plural
PS_ANIMATE	Animacy of possessor referent	Animate, Inanimate
PS_STATUS	Social status of possessor referent	Other, upperclass, God
PS_ACTIVATION	Activation (= previous mention) of possessor referent	Activated, not activated
PD_ANIMATE	Animacy of possessed referent	Animate, Inanimate
PD_ACTIVATION	Activation (= previous mention) of possessed referent	Other, upperclass, God
PD_INHERENT	Inherent vs. non-inherent possession	Inherent, non-inherent
SYNTACTIC FUNCTION	Syntactic function of the possessive phrase	Subject, object, preposition, vocative, apposition
DATIVE	Presence of a dative element in the clause	Dative, no dative
DIRECT SPEECH	Direct vs. indirect speech	Direct speech, indirect speech
NARRATIVE	Narrative vs. lyrical text passages	Narrative, lyrical

We coded for a series of features of the possessor (PS). First, we coded for person morphology (first, second, or third person) and number morphology (singular, plural) of the possessor (variables PS\_PERSON and PS\_NUMBER). Second, we coded for the animacy of the possessor referent (variable PS\_ANIMATE) (note that divine entities such as God, angels or spirits were classified as animate). Third, we coded for the social rank of the possessor, distinguishing the levels (a) other, (b) upperclass and (c) God (variable PS\_STATUS). Fourth, we coded the degree of activation of the possessor referent (PS\_ACTIVATION), giving the value ‘true’ if the possessor had been mentioned in the same or the previous sentence and ‘false’ if it had not (following the operationalization in Company Company, 2006:79).

We included three variables that characterize the possessed entity (PD). The variables PD\_ANIMATE and PD\_ACTIVATION were coded exactly as their counterparts for the description of the possessor. In order to capture the relationship of *inherent* possession



between possessor and possessed entity, we introduced the variable PD\_INHERENT. This variable received the value ‘true’ if the relationship of possession between possessor and possessed can be characterized as such that it cannot be undone by external factors (Chappell & McGregor, 1989:26-28). PD\_INHERENT received the value ‘true’ in cases such as parts of an entity (i.e., *su siniestra* ‘his left hand’), kinship terms, exuvia (blood, sweat, tears), aspects of personality including emotions, forms of personal representation (i.e., *su nombre* ‘his name’), concepts involving images of the person (i.e., *la sombra* ‘your shadow’), and important cultural concepts and objects of value (i.e., *su dios* ‘his god’, *nuestros muertos* ‘our dead’).

Two predictor variables aimed at capturing the syntactic context of the possessive construction. The first variable, SYNTACTIC FUNCTION, describes the syntactic function of the possessive phrase, with the values ‘subject’ (e.g., examples 7 or 8), ‘object’ (e.g., example 9), ‘preposition’ (when part of a prepositional phrase, as in example 10), vocative (11), and ‘apposition’ (12). Predicate objects were counted as objects. The second variable, DATIVE, received the value ‘true’ if the sentence contained a dative expression, as in example (9).

- (9) *saca-ron*                      *le*                      *luego*                      *los*                      *ojos*  
 put.out-PST.PFV.3PL    PRO.DAT.SG    right away    DET.M.PL    eyes  
 ‘and they put out his eyes’ (JU 16:21, E6/8, 13th c.)

- (10) *con*    *beso*    *de*    *la*                      *su*                      *boca*  
 with    kiss    of    DET.F.SG    POSS.3SG    mouth  
 ‘with the kisses of his mouth’ (CA 1:2, GE, 13th c.)

- (11) *levánt-a=te*                      *tú*                      *mi*                      *amiga*  
 get.up-IMP=PRO.REFL    you    POSS.1SG    friend  
 ‘Rise up, my love’ (CA 2:10, E5/7, 15th c.)

- (12) *el rey Nabucodonosor tu padre lo fiz-o*  
 DET.M.SG king Nabucodonosor POSS.2SG father him make-PST.PFV.3SG  
*mayoral*  
 master

‘[whom] the king Nebuchadnezzar thy father [...] made master’ (DAN 5:11, E5/7, 15th c.)

The last two predictor variables describe the type of text passage whether the possessive was used in direct speech (variable DIRECT SPEECH) and whether the possessive was used in a narrative or lyrical book (variable NARRATIVE).

#### *Model selection process*

After extracting and coding the data, we subjected the data to two multinomial logistic regression analyses using the function *multinom()* (Ripley & Venables, 2015) in R (R Development Core Team, 2015), one for the data from the 13th and one for the data from the 15th century. The dependent variable was the variable TYPE. We set the reference level of TYPE to ‘POSS’ (possessive adjective cases, as in *su casa* ‘his/her house’). We included each of the twelve predictor variables as predictors in the two models.<sup>4</sup>

Multinomial logistic regression analyses are not easy to interpret, as the coefficients only compare odds ratios. In order to ease interpretation we used *marginal effects* (ME), a statistical concept developed in econometrics (Cameron & Trivedi, 2010:491-492; Freese & Long, 2001:127-128). A marginal effect is the effect of a predictor variable on the dependent variable when the predictor variable is changed while the other predictor variables remain constant. In order to calculate the marginal effects, we transformed the model coefficients into the predicted probability of each of the constructional types in a certain usage context (represented by the predictor variables), fixing the covariates at a specific value. Since most

<sup>4</sup> Some of these predictor variables are not entirely independent from each other. For instance, the social rank of the possessor interacts with animacy in that both God and noblemen are always animate. Likewise, grammatical person interacts with animacy; inanimate possessors are almost never referred to using first or second person morphology. However, none of these correlations is strong enough to invalidate this analysis. The strongest correlations between the predictors are the following: PS\_Person = 1st x PS\_Animate (-0.512) and PS\_Status = God x PS\_Animate (-0.465). None of the correlations between the other predictors transcends 0.4, indicating at best a moderate correlation.

of the covariates are dummy variables (and there is no mean between TRUE and FALSE), we did not follow the usual practice to fix the covariates at their mean. Rather, we selected the level for each variable that is most frequent in our data. The corresponding fixed values are PS\_PERSON = 3rd, PS\_NUMBER = Singular, PS\_ANIMATE = Animate, PS\_STATUS = Other, PS\_ACTIVATION = Activated, PD\_ANIMATE = Inanimate, PD\_ACTIVATION = Not activated, PD\_INHERENT = Non-inherent, SYNTACTIC FUNCTION = Subject, DATIVE = No Dative, DIRECT SPEECH = Direct, NARRATIVE = Narrative. Crucially, fixing the covariates for illustration of results does not change the results of the regression analysis, which are available in the Appendix. However, it is important to keep this procedure in mind when interpreting our results in the next section. For instance, the analysis shows that ART/BARE constructions are much more likely in third person contexts whereas ART+POSS constructions are less likely. Given that we fixed the variable for person morphology to third person, the figures describing the results for the other parameters will often show that ART/BARE has a higher probability of use than ART+POSS. This would of course be different if we had fixed the variable for person morphology for, say, first person morphology. Consequently, it is important to restrict the interpretation of the figures to the *relative changes in probability* due to the influence of the variable that is being analyzed.

## RESULTS

Table 3 illustrates the overall usage frequency and the diachronic development of the four types of possessive constructions in our corpus. In both the 13th and the 15th centuries, the possessive constructions with the highest usage frequency are the POSS construction and the ART+POSS construction. They are almost evenly distributed in the 13<sup>th</sup> century, with a relative usage frequency of 44 percent (POSS) and 41 percent (ART+POSS). However, the 15th century sees a marked increase of the usage frequency of POSS, to 59 percent, at the expense of the three other constructional types, whose relative usage frequency decreases

roughly by five percentage points. Note that due to the overall higher usage frequency of ART+POSS constructions, this means that the decrease in the usage frequency is relatively weaker for ART+POSS constructions than for ART/BARE and GEN constructions. Because the usage contexts are stable in all versions of the Bible, we can exclude the possibility that this variation is due to differences in these usage contexts.

TABLE 3. *Overall usage frequency and diachronic development of possessive constructions in the Bible corpus*

	13th century (% , <i>n</i> )	15th century (% , <i>n</i> )	TOTAL
<b>POSS</b>	(44, 648)	(59, 1963)	2611
<b>ART+POSS</b>	(41, 599)	(35, 1148)	1747
<b>ART/BARE</b>	(9, 139)	(5, 168)	307
<b>GEN</b>	(6, 90)	(1, 45)	135
<b>TOTAL</b>	1476	3324	4800

*Features of the possessor entity (PS)*

Figure 1 illustrates the probability of use of the four possessive constructions according to person morphology of the possessor as predicted by the multinomial logistic regression analyses. In the 13th century (left graph), person morphology has a clear influence on the use of possessive constructions. First, the use of ART+POSS constructions is especially likely in first and second person contexts, confirming the observation by Wanner (2005:39-40). Second, the use of ART/BARE constructions is much more likely in third person contexts than in first and second person contexts (rising from 4 percent for first person and 3 percent for second person to 19 percent for third person). Third, the use of GEN constructions is basically restricted to third person contexts.

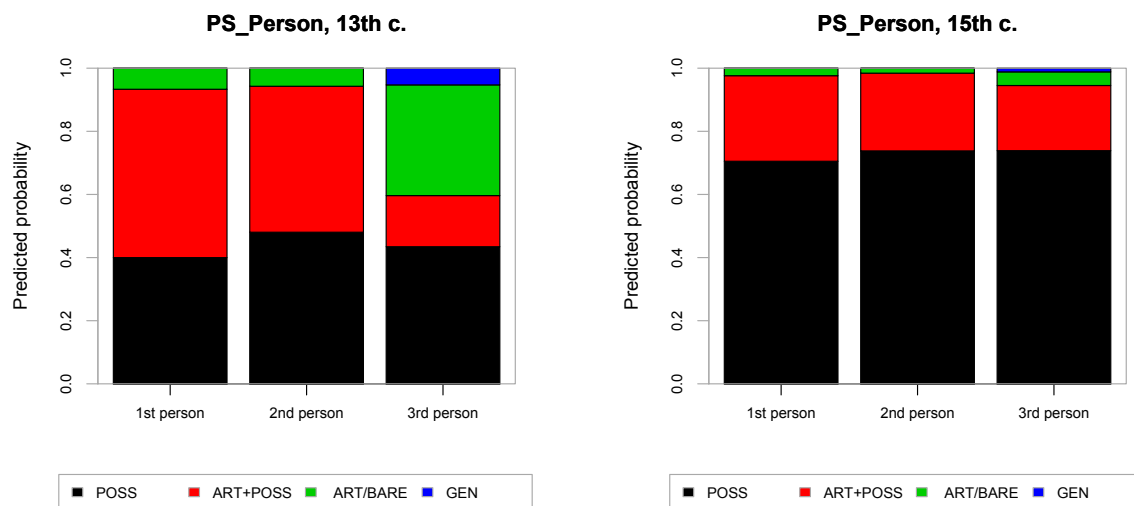


FIGURE 1. Predicted probabilities of possessive construction types by person morphology of the possessor and century.

The regression analysis on the 15th century data (right graph in Figure 1) reaches a very different result. Although a number of effects still reach statistical significance, the differences in the size of the effects are much smaller. In comparison to POSS constructions, ART+POSS constructions are still relatively more likely to appear in first and second person contexts than in third person contexts. However, this difference in the probability of use is much smaller in the 15th than in the 13th century. Note that the probability of use of ART+POSS constructions with third person morphology has actually *increased* in the 15th century. Due to the marginal effects, this explains why in many of the figures describing the results for other predictor variables, it appears that the probability of ART+POSS increases between the 13th and 15th century. However, this effect is clearly restricted to third person morphology. Likewise, the influence of person morphology on the use of the ART/BARE construction appears to be negligible in the 15th century. GEN tokens constitute an exception, as they are still restricted to third person contexts. In summary, it appears that whereas person morphology is an important predictor of the use of possessive constructions in the 13th century, by the 15th century person morphology has lost a great degree of this relevance for the use of possessive constructions.

Figure 2 illustrates the results for number morphology. In the 13th century (left graph), in comparison to POSS constructions, the use of ART+POSS is significantly more likely with singular than with plural number morphology. Inversely, the use of GEN constructions is significantly more likely with plural than singular morphology. No significant effect of number morphology on the use of ART/BARE is found. The changes in the distribution of the four constructions in the 15th century exhibit a similar pattern as the one described for person morphology. Number morphology effectively ceases to be an important predictor of the distribution of the four constructions in the 15th century, as the differences in the distribution are extremely small and do not reach the threshold of statistical significance.

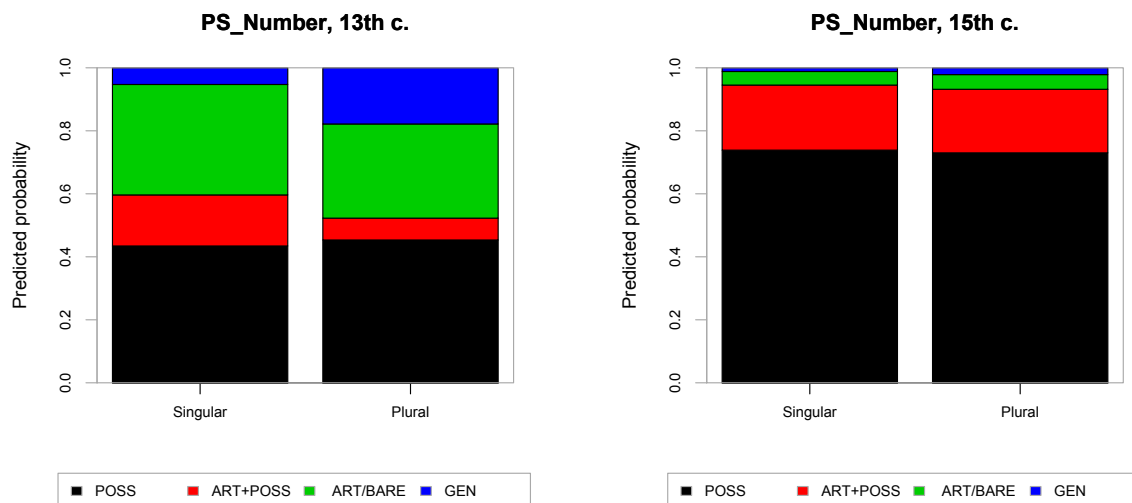


FIGURE 2. Predicted probabilities of possessive construction types by number morphology of possessor and century.

Figure 3 illustrates the results regarding possessor reference. For the 13th century, a statistically significant difference between POSS constructions and the other constructional types regarding possessor reference can be established. Whereas the use of POSS constructions is more likely with animate possessor referents, ART/BARE, ART+POSS and GEN constructions are more likely to appear with inanimate referents. This effect is especially strong for GEN constructions. Although this general pattern persists into the 15th

century, the size of these effects diminishes considerably. In addition, the effect of possessor animacy on the use of GEN loses statistical significance.

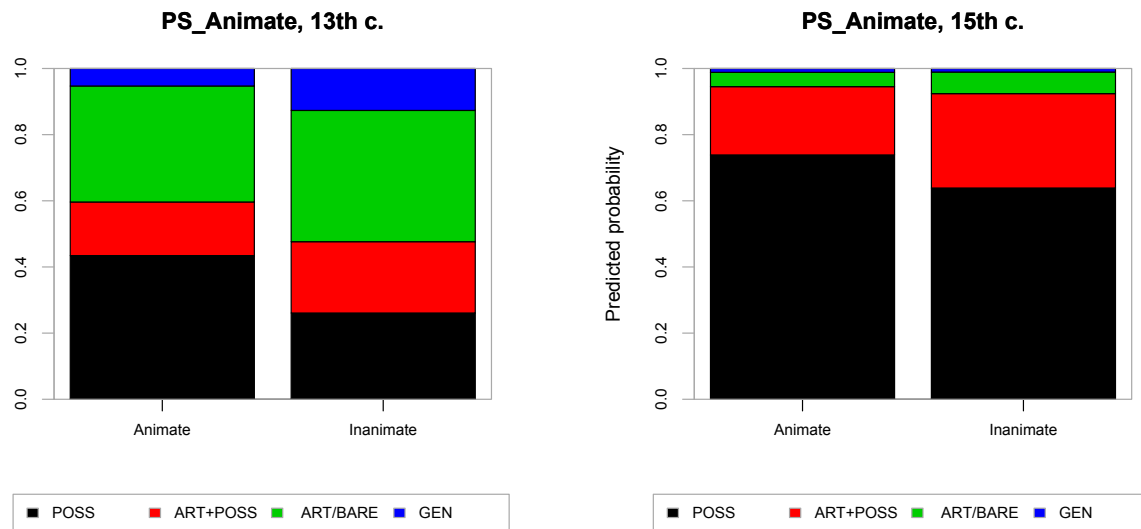


FIGURE 3. Predicted probabilities of possessive construction types by animacy and century.

Figure 4 illustrates the results from the regression analyses regarding the influence of the social status of the possessor on our four constructions. In the 13th century, in comparison to POSS, ART/BARE is significantly more likely to appear with possessor referents belonging to a lower social level (“other”) than with possessor referents belonging to the upper class or God. Although the difference between “other” and “upper class” does not reach significance for the use of ART+POSS, the analysis does show that ART+POSS is significantly more likely when the possessor referent is God than when the possessor referent belongs to a lower social level, confirming the observations by Lapesa (2000 [1970]). The use of GEN constructions is unaffected by the social status of the possessor.

The model exhibits a series of interesting changes in the 15th century. First, social status no longer has an effect on the likelihood of use of ART/BARE constructions. Second, although the impact of social status on the probability of use of ART+POSS construction has diminished, it is significantly more common not only with the possessor referent God, but also with those categorized as “upper class”. Lastly, the influence of this variable increases

for GEN constructions, which are significantly less likely with possessor referents belonging to the upper class and God than referents marked as “other”.

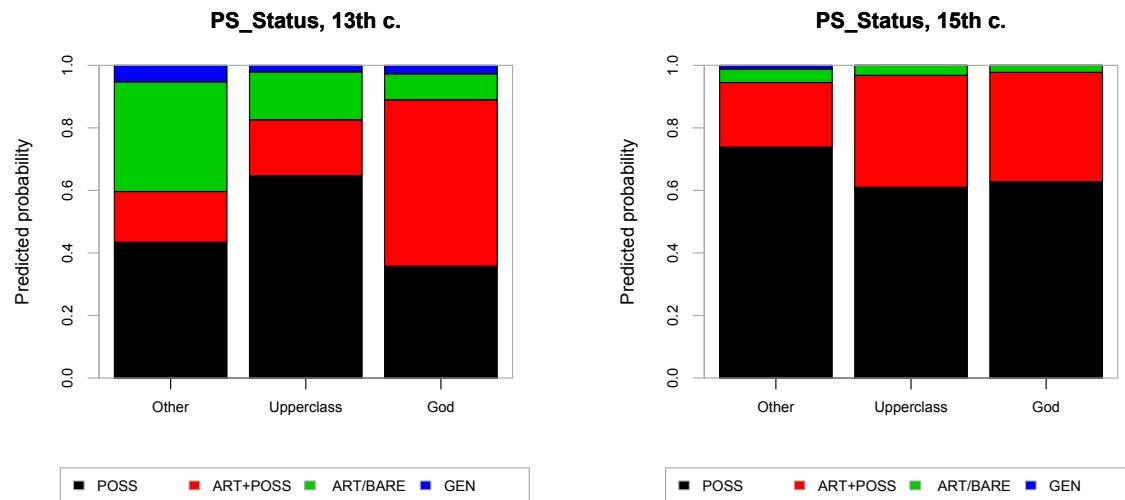


FIGURE 4. Predicted probabilities of possessive construction types by social status of possessor and century.

Finally, the variable `PS_ACTIVATION` does not have any effect on the expression of possession in the Bible translations in any century, which is why we refrain from plotting this result.

#### *Features of the possessed entity (PD)*

As illustrated by Figure 5, our statistical analysis indicates that the animacy of the possessed entity affects the choice of the possessor construction in the 13th century data. In particular, the use of both the ART/BARE construction and the GEN construction is more probable with inanimate than with animate referents. We find no statistically significant effect for ART+POSS constructions in the 13th century. There are some changes in this distribution in the 15th century. Whereas the effect for the GEN construction retains its significance, the effect for the ART/BARE construction is no longer large enough to be statistically significant. And although ART+POSS is now more likely with inanimate than with animate referents, even these statistically significant effects decrease greatly in size.



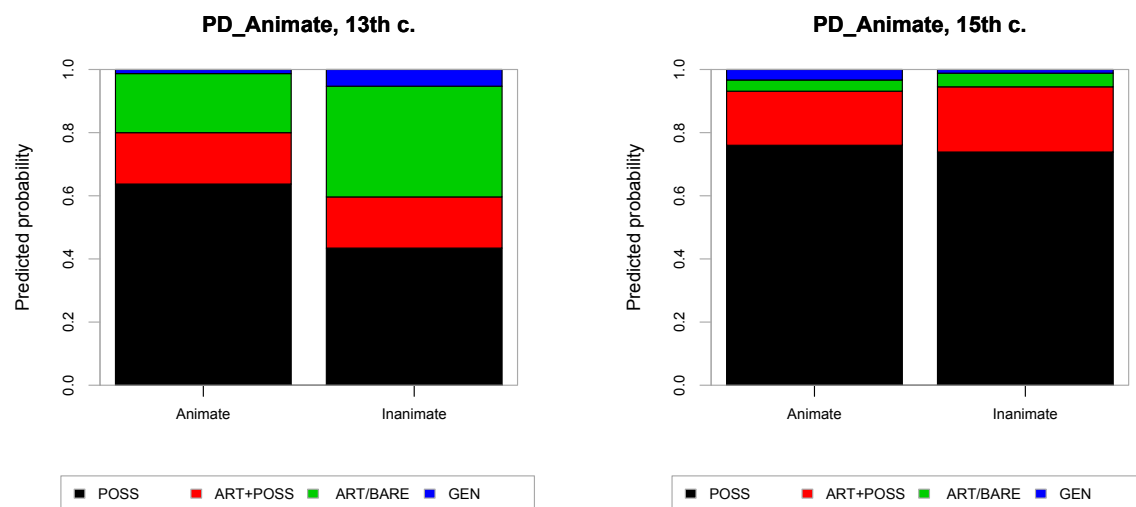


FIGURE 5. Predicted probabilities of possessive construction types by animacy of possessed entity and century.

Neither PD\_ACTIVATION nor PD\_INHERENT turn out to be good predictors of the realization of possession in our data. Whether or not the possessed entity has been mentioned earlier in the same or previous sentence (PD\_ACTIVATION) does not appear to have a significant effect in 13th century Bible translations. In contrast, in the 15th century, activation status effects point to a functional difference between ART/BARE constructions and GEN constructions. In particular, ART/BARE is significantly more likely, whereas GEN is less likely, if the possessed entity has been mentioned earlier. Regarding PD\_INHERENT, we find a significant effect in the 13th century, where inherent possession significantly favors only the use of GEN constructions. In the 15th century, no significant effects are found.

#### *Features of the syntactic context*

Turning to the syntactic function of the phrase in which the possessive is embedded (Figure 6), we find a number of interesting effects. In the 13th century ART/BARE constructions are less likely in prepositional phrases. In addition, both ART+POSS and GEN constructions are significantly more probable in syntactic phrases that are subjects. Given that subject phrases can be considered more frequent and consequently less marked syntactic configurations than,

for instance, prepositional phrases, we take this finding to hint at the smaller productivity of these constructions in comparison to the POSS construction. Once again, the right graph in Figure 6 illustrates great changes in the predicted probabilities from the 13th to the 15th century, as the POSS construction intrudes into the usage contexts of the other three constructions. For ART+POSS and GEN constructions, we observe a leveling of the effect of the syntactic context, as they appear to be relatively less restricted to the subject position. In contrast, we no longer find a negative effect of prepositional contexts on the use of ART/BARE constructions; rather, ART/BARE constructions now have a significantly elevated probability of use in vocative constructions. A closer look at the data suggests that this effect is specifically due to BARE constructions. As in many modern European languages, Old Spanish vocatives are frequently formed using a bare noun (cf. *¡Ay, hermano!* ‘Oh, brother!’).

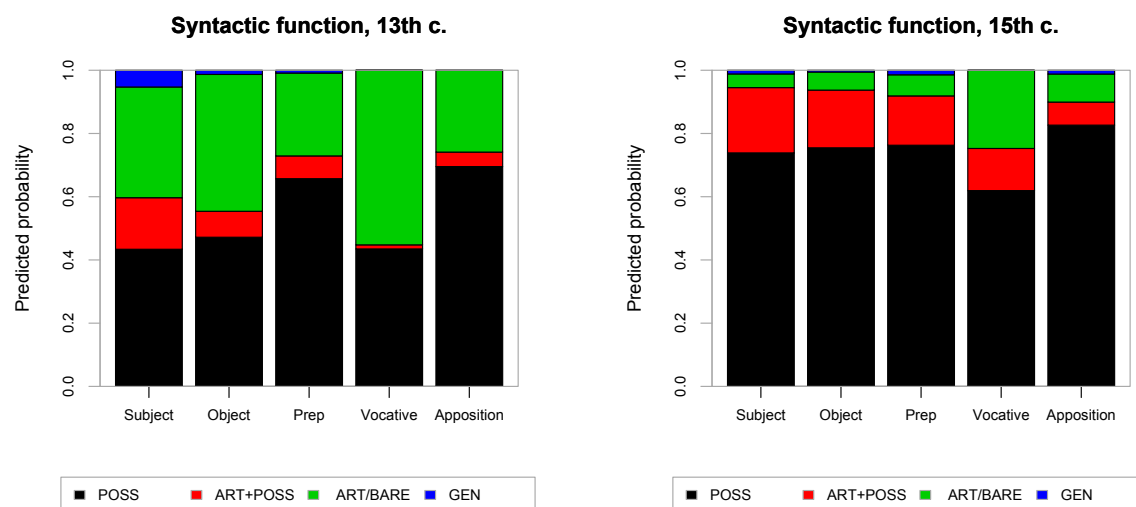


FIGURE 6. Predicted probabilities of possessive construction types by syntactic function and century.

The left graph in Figure 7 demonstrates the clear effect of the presence of a dative on the use of possessive constructions in the 13th century section of our corpus, where the use of ART/BARE is overwhelmingly more probable in sentences involving a dative (see (9)) than in sentences without a dative. This effect remains significant in the 15th century, but the

effect size decreases drastically; as illustrated in the right graph, the probability of use of POSS constructions in these contexts is now higher than .5. In addition, a significant effect of the variable Dative on the use of GEN constructions is now found; GEN constructions are significantly less probable in sentences with a dative than in ones without.

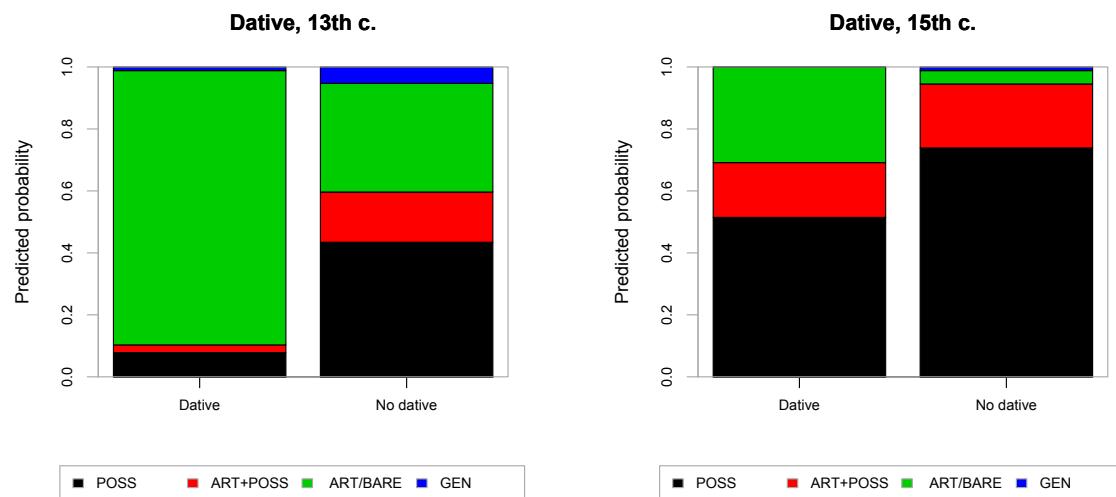


FIGURE 7. Predicted probabilities of possessive construction types by dative and century.

### *Features of text type*

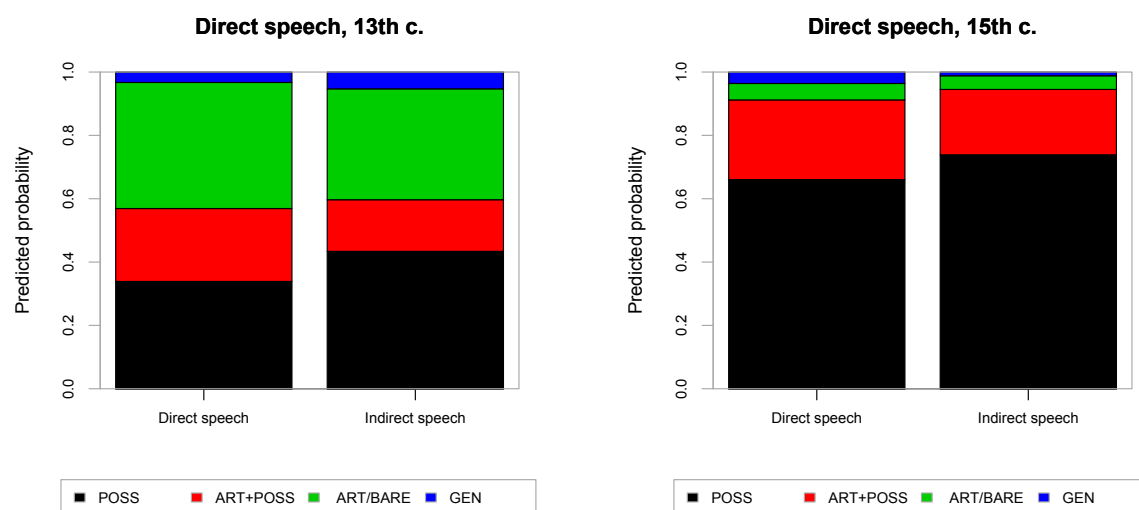


FIGURE 8. Predicted probabilities of possessive construction types by direct speech and century.

Fig. 8 illustrates the effect of whether or not the text passage in which the possessive is found is an instance of direct speech. In the 13th century, the use of an ART+POSS construction is more likely in direct speech than in non-direct speech. This effect remains stable in the 15th century. Moreover, we now also find a significant effect for GEN constructions, which are also more likely in direct speech contexts.

The last parameter evaluated is whether or not the book can be characterized as a narrative or a lyrical text. As shown in Figure 9, in the 13th century, ART/BARE and POSS constructions are more likely to be used in narrative books than in lyrical books, whereas ART+POSS and GEN constructions are more likely in lyrical books. As in the case of the variable referring to direct speech, the effect remains stable for ART+POSS constructions in the 15th century, whereas the effects for ART/BARE and GEN constructions no longer reach statistical significance.

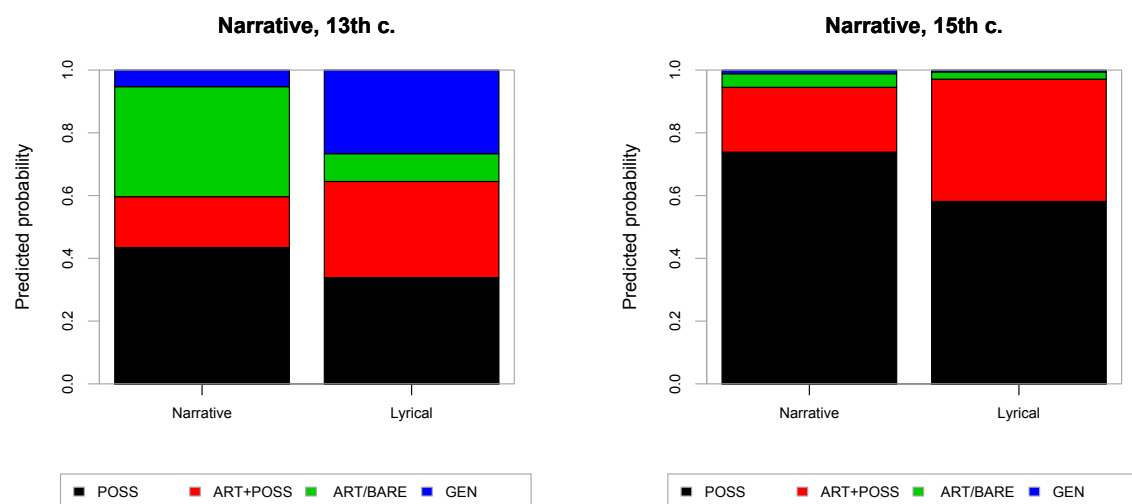


Figure 9. Predicted probabilities of possessive construction types by narrative and century.

#### DISCUSSION OF RESULTS

From the results described in the previous section, the following panorama of the expression of non-predicative possession in Old Spanish emerges. Already in the 13th century, the POSS construction (*su amigo* ‘his friend’) is the most frequent means of expression of non-predicative possession in the Bible translations; thus, in almost all usage contexts that were

investigated in this study, the POSS construction has the highest likelihood of occurrence. The probability of use of other possessive constructions can reach or transcend the probability of use of the POSS construction in very specific usage contexts. First, when the possessor referent is classified as inanimate, there is a much stronger competition between the POSS construction and the ART/BARE construction than when the possessor referent is animate. The reason appears to be that inanimate possessor referents usually imply inherent possession. For instance, in (13) the ‘possessor’ of *la solución* ‘the solution’, is *el sueño* ‘the dream’.

- (13) *Recudi-ó*                      *el*                      *rey*                      *e*                      *dixo:*                      *Bataras*                      *non*  
 answer-PST.PFV.3SG    DET.M.SG    king                      and                      say.PST.PFV.3SG    Baltasar                      not  
*te*                      *torb-e*                      *el*                      *sueño*                      *ni*                      ***la***                      ***solución***  
 PRO.DAT.SG    disturb-PRS.SUB.3SG    DET.M.SG    dream                      nor                      the                      explanation  
 ‘The king spake, and said, Belteshazzar, let not the dreame, or the interpretation thereof  
 trouble thee.’ (DAN 4:16, E6/8, 13th c.)

The parameter of inherent possession also seems to be responsible for the fact that the ART/BARE construction can compete with the POSS construction in contexts where the possessed entity is inanimate (e.g., *la solución* ‘the explanation’, in (13)). There is a correlation in our data between animacy of the possessed referent and inherent possession; inherent possession is relatively more frequent for inanimate possessed referents than for animate possessed referents ( $\chi^2=105.21(1)$ ,  $p < .001^{***}$ ). The presence of a dative pronoun constitutes a third usage context in which ART/BARE is a strong competitor to POSS. In these contexts, exemplified here in (14), ART/BARE is by far more probable than POSS. We again believe that this finding is due to the parameter of inherent possession. It is well known that in many Romance languages, the dative marks inherent possession (cf. e.g., Lamiroy & Delbecque, 1998), such as with body parts, for instance, in *sacaron le luego los ojos* ‘lit. they took him out the eyes’ (example 9 above).

In summary, although the parameter of inherent possession in itself does not reach statistical significance in the regression analysis, it nevertheless appears to play a crucial role for the use of the ART/BARE construction. This interpretation is also supported by the fact that the use of the ART/BARE construction becomes more probable when the possessed referent has been mentioned in the previous context. After all, inherent possession can be interpreted as referring to the inferability and thus, accessibility, of a referent in a given usage context.

- (14) *Taja-d       este       árbol   al       pie       e       corta-d       le*  
 cut-IMP       DET.M.SG       tree       to.the       foot       and       cut-IMP       PRO.DAT.SG  
*antes       los       ramo-s*  
 before       DET.M.PL       branch-PL

‘Hew downe the tree, and cut off his branches’ (DAN 4:11, GE, 13th c.)

As for the ART+POSS construction, the 13<sup>th</sup> century data shows that there are three usage contexts in which this structure is a serious competitor to POSS. First, if the possessor referent is God, the ART+POSS construction is more likely to be used than the POSS construction. Second, ART+POSS and POSS reach a similar likelihood of use in contexts with first and second person morphology (e.g., *la mi amiga* ‘(the) my friend’, *el to nombre* ‘(the) your name’). And third, the likelihood of use of the ART+POSS construction increases greatly in lyrical passages. In other words, the use of ART+POSS appears to be most likely exactly in those contexts in which the reference of the possessor is unambiguous. Consequently, the ART+POSS construction is not used to disambiguate a reference, but rather serves the emphatic functions of expressivity, solemnity, poeticity or reverence described by Eberenz (2000) and Lapesa 2000 (1970), among others. It is worth mentioning that the definite article has been argued to express uniqueness in a discourse universe, or in Lyon’s (1999:8) words, “the definite article signals that there is just one entity satisfying the description used.” Given that the ART+POSS construction is used precisely in those contexts

in which this uniqueness condition is not threatened, it appears that it rather serves to emphasize the uniqueness of the possessor referent.

In contrast, the results from the regression analysis do not support the view that cognitive parameters of use are crucial for the use of ART+POSS: contrary to the findings by Company Company (2006; 2009), our analysis did not identify a higher probability of use of ART+POSS in contexts where either the possessor referent or possessed referent has a high degree of mental activation via previous mention of the referent. While we do not have an explanation for this difference, our finding is supported by the results by Serradilla Castaño (2010), who also does not find an effect of activation on the selection of possessive constructions in Old Spanish.

As for the GEN construction, there is almost no usage context in which it can compete with the POSS construction. Its function appears to be the inverse of the ART+POSS in that it is mostly restricted to usage contexts in which the reference to the possessor referent or possessed entity is ambiguous, as evidenced in the finding that the GEN construction is not used in deictic contexts (first and second person morphology) and is less likely in inherent possession contexts. This ambiguity-resolving function is evident in examples such as (15).

- (15) *Meti-ó                    la                    hueste   su                    mano   a   toda-s   las*  
 put-PST.PFV.3SG   DET.F.SG   army   POSS.3SG   hand   to   all-PL   DET.F.PL  
*cosa-s   d'ella                    que   de   dessear   eran*  
 thing-PL   of.DET.F.SG   that   of   desire   be.PST.IPFV.3PL  
 ‘The adversary [lit. army] hath spread out his hand upon all her pleasant things’ (DAN 4:11, GE, 13th c.)

The possessor referent of *las cosas d'ella* ‘the things of her’ is the city of Jerusalem, not the subject of the sentence, *la hueste* ‘the (feminine) army’. Consequently, the use of the GEN construction in example (15) appears to indicate to the reader that the possessor referent is “unexpected” in the sense that it cannot be inferred from the context. This assumption is also

supported by the finding that at least in the 15th century data the use of the GEN construction is less likely (a) when the possessed referent has been mentioned in the previous context and (b) in contexts involving a dative.

To some degree, GEN also appears to have a stylistic function. In our 13th century data, its probability of use is much higher in lyrical text passages than in narrative text passages, confirming the previous analysis by Enrique-Arias (2012a). Referential ambiguity is clearly not an issue in examples such as (16) where the possessor referent (the garden) is present in the same sentence as the possessive phrase and no other possessor referent is available (if the possessive phrase were to refer to the wind, we would expect second person morphology). Rather, the GEN construction appears to be used for stylistic reasons, maybe to avoid a repetition of the ART+POSS construction used in the same sentence.

- (16) *Levant-a=t                    sierço                    e                    uen                    abrego.                    solla                    el*  
 rise-IMP=PRO.REFL    north wind    and    come.IMP    south wind    blow.IMP    DET.M.SG  
*mio                    uerto                    e                    correr-an                    los                    unguento-s                    d=el*  
 POSS.1.SG    garden    and    run-FUT.3PL    DET.M.PL    ointment-PL    of=DET.M.SG  
 ‘Awake, O north wind; and come, thou south; blow upon my garden, *that* the spices thereof  
 may flow out’ (CA 4:16, E6/8, 13th c.)

The analysis has shown that in contexts such as (16), the use of GEN constructions is almost as likely as the use of POSS and ART+POSS constructions. However, it has to be noted that this increase in the likelihood of GEN constructions is only relevant for sentences in which the possessive phrase displays third person morphology.

The distribution of possessive constructions described in this section undergoes a series of important changes between the 13th and 15th century Bible translations. First, the use of the POSS construction rises dramatically to become the dominant possessive construction in virtually all contexts. This extension of the POSS construction leads to a leveling of the influence of predictor variables on the use of the other possessive constructions. For instance,



whereas person and number morphology were important predictors for the use of ART+POSS in the 13th century, these variables no longer have a significant influence in the 15th century. Likewise, although the parameters of social status of the possessor, type of possessed referent and narrative text type have a significant influence on the use of ART/BARE in the 13th century, these effects disappear in the 15th century analysis. At the same time the POSS construction gradually intrudes into contexts involving a dative construction, in which the use of the ART/BARE construction is overwhelmingly frequent in the 13th century. In the same vein, whereas inherent possession and number morphology are important predictors for the use of the GEN construction in the 13th century (the use of GEN is more frequent with inherent possession and plural morphology), these effects are no longer significant in the 15th century data.

These developments can be described as an instance of diffusion (De Smet, 2012) or possibly *capitalization* (Pountain, 2000). By extending its usage frequency in other contexts, the POSS construction gradually takes over functions previously associated with other possessive constructions, creating the leveling effect.

There are only a few usage contexts that constitute an exception to this general trend. In regards to the ART+POSS construction, while its overall frequency decreases dramatically in the 15th century translations, its use continues, or even increases, in a number of contexts that can be described as emphatic: first, when the possessor referent is classified as belonging to the upper class (such as God, the king or noblemen); second, when the NP that contains the possessive structure functions as a vocative; and third, in lyrical (vs. narrative) passages and direct (vs. indirect) speech. This observation conforms to recent models of actualization processes which assume that usage contexts that are more typical for the use of a construction are affected later by replacement processes while less typical usage contexts tend to adopt the new structure first (an effect termed 'remanence' in Rosemeyer, 2014:89-90). This increasing restriction of ART+POSS to emphatic uses promotes an interpretation of the alternation

between POSS and ART+POSS in terms of stylistic factors, thus explaining why ART+POSS might even experience a relative strengthening in these contexts.

#### CONCLUSION

In this paper we have demonstrated that the coupling of data from the parallel Bible translation corpus and multinomial logistic regression analysis can help overcome two key problems in quantitative analyses of syntactic change in general, and that it can also significantly improve the methodological rigor of variationist approaches to syntactic change. Regarding the problem of the definition of the variable context, the use of a parallel corpus such as *Biblia Medieval* allows for direct comparison of linguistic variants. Consequently, it was ensured that according to the linguistic knowledge of the translators of the Bible, each of the tokens in the corpus—even the structures with no explicit possessive markers such as ART/BARE—actually expresses possession. The use of the *Biblia Medieval* corpus also alleviates the problem of the comparability of contexts because the contexts of occurrence of the structures under scrutiny do not change over time. Consequently, the changes in the distribution of the possessive constructions demonstrated in our analysis (such as the successive replacement of ART+POSS with POSS) are not likely to be due to differences in genres, registers, styles or the contents of the texts in the corpus; rather, they reflect actual syntactic changes. A multinomial regression is the correct analytical tool for such data. Specifically, by using multinomial regression analyses with marginal effects for our data, we were able to estimate for each variant whether or not it represents a serious competitor to another variant in a specific context. This procedure yields both a highly reliable synchronic description of the function of the different variants and of the diachronic trajectory of these variants.

Applying this methodology to the case of non-predicative possession in Old Spanish has allowed for an in-depth analysis of the factors governing the selection of the different types of

possessive constructions, as well as their diachronic development. Our analysis has shown that already in the 13th century data, the POSS construction (*su amigo* ‘her friend’) was the dominant construction in a majority of the investigated usage contexts. Only in some usage contexts is it challenged by the other possessive constructions. First, the probability of ART/BARE constructions (such as *el amigo* ‘her friend’) is especially high in contexts that relate to the parameter of inherent possession. In particular, the ART/BARE construction can compete with the POSS construction in contexts involving a dative. In such cases the use of an explicit possessive marker appears not to be relevant because of the overall high accessibility of the possessed referent. Second, our data suggests that the apparent functional redundancy of the ART+POSS construction (definite determiner + definite possessive pronoun) serves an emphatic function. The use of the ART+POSS construction is equally or more likely than the POSS construction in contexts that serve to express pragmatic functions such as reverence, solemnity or poeticality. Thus, the ART+POSS construction is used to emphasize the importance of the possessor referent, which is why it is overrepresented in sentences in which the possessor is either the speaker or the addressee, or God. Likewise, the ART+POSS construction is a much more serious competitor to the POSS construction in lyrical text passages than in narrative text passages. Third, the GEN construction typically has an ambiguity-resolving function, as evinced by the fact that it is most frequent in usage contexts in which the reference to the possessor referent or possessed entity is ambiguous. However, our analysis has also demonstrated the relevance of stylistic issues, as the GEN construction can compete with POSS and ART+POSS constructions in lyrical third-person contexts. In summary, our methodology has allowed us to identify to which degree the four possessive constructions compete in a given usage context and has consequently led to a precise characterization of the function of each of these constructions.

Likewise, the analysis has enabled us to establish a description of the changes in the distribution of possessive constructions between the 13th and 15th century that is unaffected

by contextual variation. Our results suggest an extension of the usage of POSS constructions between the 13th and the 15th century in terms not only of usage frequency, but also of probability of use in specific usage contexts. Consequently, POSS constructions start displacing the other types of possessive constructions even in contexts in which these had been strong competitors. For instance, POSS constructions massively intrude into dative sentences in the 15th century, displacing ART/BARE constructions, as well as first and second person contexts, displacing ART+POSS constructions. These changes have a leveling effect, obliterating most of the functional differences between the four constructions identified in our 13th century data. Importantly however, the analysis has also illustrated some exceptions to this general trend. In particular, there is a gradual strengthening of the opposition between POSS and ART+POSS constructions in terms of stylistic parameters. Whereas most of the structural parameters that characterize the opposition in the 13th century have a much lower incidence in the 15th century data (e.g., person and number morphology, as well as animacy of possessor and possessed referent), stylistic parameters—in particular, the social status of the possessor, direct speech and narrative/lyrical text type—continue to exercise an important influence on the opposition in the 15th century data. In some cases, the relevance of these parameters even increases in the 15th century. The obsolescence of ART+POSS constructions and the subsequent loss of functional oppositions led to a refunctionalization of the opposition between POSS and ART+POSS in terms of stylistic parameters, with ART+POSS as the stylistically marked variant. Our study thus illustrates how in reductive processes of language change, variation due to structural constraints yields to stylistic variation.

## APPENDIX

Tables 7 and 8 illustrate the results from the multinomial regression analyses over the 13th century and 17th century data. They give the coefficient (COEFF) and the *p*-value (P) calculated for each of the levels of the predictor variables for each of the four levels of the dependent variable TYPE. Since the reference level of the dependent variable is set to POSS, the coefficients refer to the probability of use of one of the three other levels (ART+POSS, ART/BARE, or GEN) in comparison to POSS in these specific usage contexts. For instance, the coefficients in the fourth line in Table 7 indicate that if the possessive construction has first person morphology (PS\_PERSON = 1st) instead of third person morphology (PS\_PERSON = 3rd), the likelihood of use of ART + POSS in comparison to the use of POSS increases by 1.274, whereas the likelihood of use of ART/BARE decreases by 1.576. The *p*-values indicate that these effects are highly significant ( $p < .001^{***}$ ). Note that for GEN, no *p*-value is given for the effect of person morphology because we only find GEN in third person contexts in Old Spanish (in the 15th century, we do find some tokens of GEN in first and second person contexts, as indicated by Table 8). The tables also give the total number of occurrences for each variable level, as well as the relative frequencies of the four constructions for each variable level.

TABLE 7. Multinomial logistic regression analysis, 13th century (reference level of the dependent variable = POSS)

VARIABLE	LEVEL	n	POSS	ART+POSS			ART/BARE			GEN		
			REL. FREQ. %	REL. FREQ. %	COEFF	P	REL. FREQ. %	COEFF	P	REL. FREQ. %	COEFF	P
PS_Person	3rd	667	46	23	Reference level		17	Reference level		14	Reference level	
	1st	518	43	54	1.274	0.000	3	-1.576	0.000	0	-92.674	NA
	2nd	291	39	58	0.948	0.000	3	-1.912	0.000	0	-94.390	NA
PS_Number	Singular	1259	43	43	Reference level		9	Reference level		5	Reference level	
	Plural	217	51	24	-0.892	0.000	11	-0.202	0.509	14	1.172	0.008
PS_Animate	FALSE	187	28	36	Reference level		16	Reference level		20	Reference level	
	TRUE	1289	46	41	-0.796	0.001	9	-0.634	0.055	4	-1.384	0.001
PS_Status	Other	1134	41	42	Reference level		10	Reference level		7	Reference level	
	Upperclass	239	64	28	-0.299	0.182	7	-1.227	0.001	1	-1.320	0.107
	God	103	33	59	1.383	0.000	3	-1.247	0.052	5	-0.475	0.441
PS_Activation	FALSE	115	42	24	Reference level		14	Reference level		20	Reference level	
	TRUE	1361	44	42	0.055	0.852	9	0.210	0.567	5	-0.224	0.548
PD_Animate	FALSE	1001	38	43	Reference level		11	Reference level		8	Reference level	
	TRUE	475	56	35	-0.379	0.018	7	-1.011	0.000	2	-1.790	0.000

<b>PD_Activation</b>	FALSE	1223	42	41	<i>Reference level</i>		10	<i>Reference level</i>		7	<i>Reference level</i>	
	TRUE	253	51	38	-0.139	0.447	8	-0.101	0.742	3	0.273	0.553
<b>PD_Inherent</b>	FALSE	377	52	36	<i>Reference level</i>		9	<i>Reference level</i>		3	<i>Reference level</i>	
	TRUE	1099	41	42	0.222	0.157	10	0.021	0.934	7	0.840	0.039
<b>Syntactic function</b>	Subject	517	30	55	<i>Reference level</i>		8	<i>Reference level</i>		7	<i>Reference level</i>	
	Object	408	42	36	-0.767	0.000	14	0.128	0.653	8	-1.469	0.000
	Prep	476	56	32	-1.227	0.000	8	-0.705	0.015	4	-2.192	0.000
	Vocative	55	73	24	-2.537	0.000	3	0.450	0.593	0	-7.155	0.000
	Apposition	20	80	15	-1.738	0.014	5	-0.774	0.480	0	-224.547	0.000
<b>Dative</b>	FALSE	1399	45	41	<i>Reference level</i>		8	<i>Reference level</i>		6	<i>Reference level</i>	
	TRUE	76	22	28	-0.160	0.659	47	2.642	0.000	3	0.208	0.808
<b>Direct speech</b>	FALSE	752	48	33	<i>Reference level</i>		11	<i>Reference level</i>		8	<i>Reference level</i>	
	TRUE	724	40	48	0.590	0.000	8	0.375	0.151	4	-0.232	0.526
<b>Narrative</b>	FALSE	890	35	52	<i>Reference level</i>		4	<i>Reference level</i>		9	<i>Reference level</i>	
	TRUE	586	57	24	-0.885	0.000	17	1.128	0.000	2	-1.865	0.000

TABLE 8. Multinomial logistic regression analysis, 15th century (reference level of the dependent variable = POSS)

VARIABLE	LEVEL	n	POSS	ART+POSS			ART/BARE			GEN		
			REL. FREQ. %	REL. FREQ. %	COEFF	P	REL. FREQ. %	COEFF	P	REL. FREQ. %	COEFF	P
PS_Person	3rd	1663	62	29	Reference level		7	Reference level		2	Reference level	
	1st	1057	56	40	0.319	0.005	4	-0.550	0.038	0	-16.058	0.000
	2nd	604	57	41	0.177	0.147	2	-1.042	0.003	0	-3.083	0.003
PS_Number	Singular	2862	59	35	Reference level		5	Reference level		1	Reference level	
	Plural	462	63	31	-0.011	0.927	4	0.090	0.740	2	0.622	0.149
PS_Animate	FALSE	447	52	39	Reference level		7	Reference level		2	Reference level	
	TRUE	2877	60	34	-0.469	0.000	5	-0.556	0.040	1	-0.069	0.878
PS_Status	Other	2599	60	33	Reference level		5	Reference level		2	Reference level	
	Upperclass	512	58	37	0.741	0.000	5	-0.111	0.688	0	-15.756	0.000
	God	213	52	46	0.689	0.000	2	-0.496	0.317	0	-14.520	0.000
PS_Activation	FALSE	301	64	27	Reference level		6	Reference level		3	Reference level	
	TRUE	3023	59	35	0.133	0.391	5	0.217	0.468	1	-0.439	0.283
PD_Animate	FALSE	2303	57	37	Reference level		5	Reference level		1	Reference level	
	TRUE	1021	63	29	-0.216	0.029	6	-0.227	0.294	2	1.019	0.005



<b>PD_Activation</b>	FALSE	2693	58	36	<i>Reference level</i>		4	<i>Reference level</i>		2	<i>Reference level</i>	
	TRUE	631	62	28	-0.144	0.190	9	0.552	0.005	1	-1.665	0.007
<b>PD_Inherent</b>	FALSE	786	59	36	<i>Reference level</i>		4	<i>Reference level</i>		1	<i>Reference level</i>	
	TRUE	2538	59	34	-0.067	0.483	5	0.078	0.723	2	0.715	0.202
<b>Syntactic Function</b>	Subject	1198	55	40	<i>Reference level</i>		4	<i>Reference level</i>		1	<i>Reference level</i>	
	Object	858	59	34	-0.145	0.145	6	0.259	0.266	1	-0.738	0.137
	Prep	1111	63	30	-0.313	0.001	5	0.403	0.073	2	0.178	0.622
	Vocative	106	52	35	-0.261	0.287	13	1.921	0.000	0	-3.540	0.000
	Apposition	51	78	12	-1.154	0.012	8	0.608	0.292	2	-0.087	0.938
<b>Dative</b>	FALSE	3201	60	35	<i>Reference level</i>		4	<i>Reference level</i>		1	<i>Reference level</i>	
	TRUE	123	43	33	0.207	0.353	24	2.329	0.000	0	-11.470	0.000
<b>Direct speech</b>	FALSE	1735	62	32	<i>Reference level</i>		5	<i>Reference level</i>		1	<i>Reference level</i>	
	TRUE	1589	56	37	0.310	0.000	5	0.296	0.139	2	1.237	0.000
<b>Narrative</b>	FALSE	1817	53	43	<i>Reference level</i>		4	<i>Reference level</i>		0	<i>Reference level</i>	
	TRUE	1507	67	24	-0.874	0.000	7	0.395	0.056	2	0.427	0.235

## REFERENCES

- Cameron, A. Colin, & Trivedi, Pravin K. (2010). *Microeconometrics using Stata*, 2nd edition. Texas: Stata Press.
- Chappell, Hilary, & McGregor, William B. (1989). Alienability, inalienability and nominal classification. *Berkeley Linguistics Society Proceedings* 15:24–36.
- Company Company, Concepción. (1994). Semántica y sintaxis de los posesivos duplicados en el español de los siglos XV y XVI. *Romance Philology* 48(2):111-135.
- Company Company, Concepción. (2006). Persistencia referencial, accesibilidad y tópico. La semántica de la construcción artículo + posesivo + sustantivo en el español medieval. *Revista de Filología Española* 86(1):65-103.
- \_\_\_\_\_. (2009). Artículo + posesivo + sustantivo y estructuras afines. In C. Company Company (ed.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal*. México: Fondo de Cultura Económica & Universidad Nacional Autónoma de México. 759-880.
- De Vries, Lourens. (2007). Some remarks on the use of Bible translations as parallel texts in linguistic research. In M. Cysow & B. Wälchli (eds.), *Parallel Texts: Using translational equivalents in linguistic typology*. Special issue of *Sprachtypologie und Universalienforschung (STUF)* 60:95-99.
- De Smet, Hendrik. (2012). The course of actualization. *Language* 88(3):601-633.
- Eberenz, Rolf. (2000). *El español en el otoño de la Edad Media. Sobre el artículo y los pronombres*. Madrid: Gredos.
- Enrique-Arias, Andrés. (2008). Biblias romanceadas e historia de la lengua. In C. Company Company & J. G. Moreno de Alba (eds.), *Actas del VII Congreso Internacional de Historia de la Lengua Española*. Madrid: Arco Libros. 1781-1794.

- \_\_\_\_\_. (2009). Ventajas e inconvenientes del uso de Biblia Medieval (un corpus paralelo y alineado de textos bíblicos) para la investigación en lingüística histórica del español. In A. Enrique-Arias (ed.), *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. Frankfurt/Madrid: Vervuert/Iberoamericana. 269-283.
- \_\_\_\_\_. (2012a). *Lingua eorum – la lengua d'ellos*: sobre la suerte de un calco sintáctico del latín en la historia del español. *Bulletin of Hispanic Studies* 89:813-829.
- \_\_\_\_\_. (2012b). Dos problemas en el uso de corpus diacrónicos del español: perspectiva y comparabilidad. *Scriptum digital* 1:85-106.
- \_\_\_\_\_. (2013). On the usefulness of using parallel texts in diachronic investigations: insights from a parallel corpus of Spanish medieval Bible translations. In P. Bennett, M. Durrell, S. Scheible, & R. J. Whitt (eds.), *New Methods in Historical Corpora*. Tübingen: Gunter Narr. 105-115.
- Enrique-Arias, Andrés, & Pueyo Mena, F. Javier. (2008-2016). *Biblia Medieval*. Available online at <http://www.bibliamedieval.es>. Last access 26 November 2014.
- Freese, Jeremy, & Long, J. Scott. (2001). *Regression models for categorical dependent variables using Stata*. College Station, Tex.: Stata Corporation.
- Kaiser, Georg A. (2005). Bibelübersetzungen als Grundlage für empirische Sprachwandeluntersuchungen. In C. D. Pusch, J. Kabatek & W. Raible (Eds.), *Romance Corpus Linguistics II. Corpora and Diachronic Linguistics*. Tübingen: Gunter Narr. 71-83.
- Krawczak, Karolina. (2014). Shame and its near-synonyms in English: A multivariate corpus-driven approach to social emotions. In I. Novakova, P. Blumenthal, & D. Siepmann (eds.), *Emotions in Discourse*. Frankfurt/Main: Peter Lang. 84-94.
- Labov, William. (1982). Building on empirical foundations. In W. P. Lehmann & Y. Malkiel (eds.), *Perspectives on Historical Linguistics*. Amsterdam, Philadelphia: John Benjamins. 17-92.

- Lapesa, Rafael. (2000 [1970]). Sobre el artículo ante posesivo en castellano antiguo. In R. Cano Aguilar, & M. T. Echenique Elizondo (eds.), *Estudios de morfosintaxis histórica del español*. Madrid: Gredos. 413-435.
- Lamiroy, Béatrice, & Delbecque, Nicole. (1998). The possessive dative in Romance and Germanic languages. In W. Van Langendonck & W. Van Belle (eds.), *The Dative. Volume 2: Theoretical and contrastive studies*. Amsterdam, Philadelphia: John Benjamins. 29-74.
- Li, Xiaojian, Schweickert, Richard, & Gandour, Jack. (2000). The phonological similarity effect in immediate recall: Positions of shared phonemes. *Memory & Cognition* 28(7): 1116-1125.
- Lyons, Christopher. (1999). *Definiteness*. Cambridge: Cambridge University Press.
- McEnery, Tony, & Xiao, Zhonghua. (2007). Parallel and comparable corpora: The state of play. In Y. Kawaguchi, T. Takagaki, N. Tomimori, & Y. Tsuruga (eds.), *Corpus-Based Perspectives in Linguistics*. Amsterdam, Philadelphia: John Benjamins. 231-142.
- Orme, John G., & Combs-Orme, Terri, (2009). *Multiple Regression with Discrete Dependent Variables*. Oxford: Oxford University Press.
- Pountain, Christopher J. (2000). Capitalization. In J. C. Smith & D. Bentley (eds.), *Historical Linguistics 1995. Volume 1: General Issues and non-Germanic Languages*. Amsterdam, Philadelphia: John Benjamins. 295-309.
- R Development Core Team (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at <http://www.R-project.org>. Last access 21 September 2015.
- Ratcliff, Peter, & McKoon, Gail. (2001). A multinomial model for short-term priming in word identification. *Psychological Review* 108(4):835-846.

- Ripley, Brian, & Venables, William. (2015). *nnet. Software for feed-forward neural networks with a single hidden layer, and for multinomial log-linear models*. R package version 7-3-11.
- Rosemeyer, Malte. (2014). *Auxiliary Selection in Spanish. Gradience, Gradualness, and Conservation*. Amsterdam, Philadelphia: John Benjamins.
- Sankoff, David. & Thibault, Pierrette. (1981). Weak complementarity: tense and aspect in Montreal French. In B. B. Johns & D. R. Strong (eds.), *Syntactic Change*. Ann Arbor: University of Michigan. 205-216.
- Serradilla Castaño, Ana. (2010). "Artículo + posesivo + nombre" frente a "posesivo + nombre" como *variante invisible* en un texto medieval. *Epos XXVI*. 53-76.
- Tagliamonte, Sali. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. (2012). *Variationist Sociolinguistics. Change, Observation, Interpretation*. Malden, MA: Wiley-Blackwell.
- Wanner, Dieter. (2005). The corpus as a key to diachronic explanation. In J. Kabatek, C. D. Pusch, & W. Raible (eds.), *Romance Corpus Linguistics II. Corpora and Diachronic Linguistics*. Tübingen: Gunter Narr. 31-44.